Data Adaptive Rank-Shaping Methods for Solving Least Squares Problems

Anthony J. Thorpe and Louis L. Scharf, *Fellow, IEEE*

***Abstract*— There are two types of problems in the theory of least squares signal processing: parameter estimation and signal extraction. Parameter estimation is called "inversion" and signal extraction is called "filtering." In this paper, we present a unified theory of rank shaping for solving overdetermined and underdetermined versions of these problems. We develop several data-dependent rank-shaping methods and evaluate their performance. Our key result is a data-adaptive Wiener filter that automatically adjusts its gains to accommodate realizations that are *a priori* unlikely. The adaptive filter dramatically outperforms the Wiener filter on atypical realizations and just slightly underperforms it on typical realizations. This is the most one can hope for in a data-adaptive filter.**

I. INTRODUCTION

THE principle of least squares is to fit a model to a set of observations in such a way as to minimize the squares of the errors between the observations and the model—hence the term *least squares*. Rank shaping is a general method for reducing the variance of an estimator at the expense of introducing model bias. In doing this, we hope to reduce the mean-squared error (MSE), which is the sum of variance and squared bias.

We will examine rank shaping in overdetermined and underdetermined least squares problems. In the overdetermined problem, we fit a simple model to a large, complex data set, while in the underdetermined problem we fit a complex model to a small, simple data set. We develop several data-dependent procedures for shaping the rank of least squares estimators. Our most promising solution is a mix between rank-shaped least squares and data-adaptive Wiener filtering. In this solution, a prior distribution is assigned to the parameter of interest, and this distribution is used to assign a prior distribution to the rank-shaping gain one would use in a least squares solution. Then, the measured data is used to compute the conditional mean of this gain. This conditional mean is, in fact, the data-adaptive gain of the adaptive Wiener filter. The filter has very high performance on unlikely data and nearly Wiener performance on likely data. Our methods are similar in spirit to the shrinkage methods of James and Stein [1], Marquardt [2], and Stein [3]; however, our data adaptive shrinkage takes place mode-by-mode.

Manuscript received August 17, 1993; revised December 19, 1994. Supported by the Office of Naval Research, Mathematics Division, under contract No. N00014-89-J-1070 and by Bonneville Power Administration under Contract #DEBI7990BPO7346. The associate editor coordinating the review of this paper and approving it for publication was Dr. Barry Sullivan.

A. Thorpe is with Analytical Surveys, Inc., Colorado Springs, CO 80901 USA.

L. Scharf is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309 USA.

IEEE Log Number 9412007.

Our philosophy in this paper is that with clairvoyant side information (which we do not have), we could improve on least squares for estimating signals and parameters. A natural inclination is then to try to steal this clairvoyant information from the data. We show that this is extremely risky, that naive methods cannot work, and that only sophisticated, conservative deviations from Wiener filtering can work. The result is a nonlinear filter that uses mode-dependent, nonlinear companders to estimate something akin to Wiener gain.

A. The Linear Statistical Model

The linear statistical model is a signal-plus-noise model: the observations consist of a model or signal component and an error or noise component. Moreover, the signal component satisfies a set of linear equations. This leads to the model

$$\mathbf{y} = \mathbf{x} + \mathbf{n}; \qquad \mathbf{x} = \mathbf{H}\underline{\boldsymbol{\theta}} \tag{1.1}$$

where $\mathbf{y}$ is a noisy $N \times 1$ observation of the signal $\mathbf{x}$. The matrix $\mathbf{H}$ is the $N \times p$ model matrix, and $\underline{\boldsymbol{\theta}}$ is the $p \times 1$ parameter vector. Geometrically, the signal $\mathbf{x}$ lies in the rank-p subspace $\langle \mathbf{H} \rangle$, illustrated in Fig. 1. The signal $\mathbf{x}$ can be thought of as a linear combination of columns of $\mathbf{H}$

$$\mathbf{x} = \sum_i \mathbf{h}_i \theta_i; \qquad \mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_p]. \tag{1.2}$$

Each $\mathbf{h}_i$ might be a mode in a system. We wish to determine the weights θ_i. Alternatively, the observation $\mathbf{y}$ could be a noisy version of some modulated information $\underline{\boldsymbol{\theta}}$ that we are trying to estimate. The linear model also arises in curve-fitting problems such as polynomial interpolation.

We will make extensive use of the singular value decomposition of $\mathbf{H}$, namely $\mathbf{H} = \mathbf{U}\underline{\boldsymbol{\Sigma}}\mathbf{V}^T$, where $\underline{\boldsymbol{\Sigma}}$ is the diagonal matrix of singular values σ_i. In the overdetermined case, where $N > p$, $\mathbf{U}$ is $N \times p$, $\underline{\boldsymbol{\Sigma}}$ is $p \times p$, and $\mathbf{V}^T$ is $p \times p$. In the underdetermined case, where $p > N$, $\mathbf{U}$ is $N \times N$, $\underline{\boldsymbol{\Sigma}}$ is $N \times N$, and $\mathbf{V}^T$ is $N \times p$. Note that, in the overdetermined case, we have $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}$, but in the underdetermined case we have $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}$. These matrix decompositions are illustrated below

$$\text{i) overdetermined} \qquad \begin{bmatrix} \mathbf{H} \\ \\ \end{bmatrix} = \begin{bmatrix} \mathbf{U} \\ \\ \end{bmatrix} [\underline{\boldsymbol{\Sigma}}][\mathbf{V}^T]$$

$$\text{ii) underdetermined} \qquad [\mathbf{H} \quad\quad] = [\mathbf{U}][\underline{\boldsymbol{\Sigma}}][\mathbf{V}^T \quad\quad]. \tag{1.3}$$

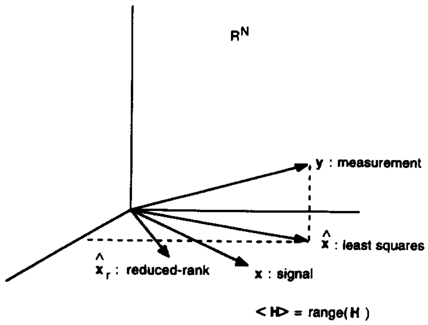


Fig. 1. Pictorial interpretation of the overdetermined least squares filtering problem.

B. Distributions

Throughout our analysis, we assume that the model matrix $\mathbf{H}$ is known and that the noise $\mathbf{n}$ is distributed as $N[0, \sigma^2 \mathbf{I}]$ with σ^2 also known. When the vectors $\underline{\boldsymbol{\theta}}$ and $\mathbf{x}$ are deterministic, then $\mathbf{y}$ is distributed as $N[\mathbf{x}, \sigma^2 \mathbf{I}]$. When $\underline{\boldsymbol{\theta}}$ and $\mathbf{x}$ are random, then $\mathbf{y}$ is conditionally distributed as $N[\mathbf{x}, \sigma^2 \mathbf{I}]$.

II. Rank-Shaped Filtering and Inversion

We consider two problems: filtering and inversion. For the inversion problem, the least squares estimate $\hat{\underline{\boldsymbol{\theta}}}$ of the parameter vector $\underline{\boldsymbol{\theta}}$ is $\hat{\underline{\boldsymbol{\theta}}} = \mathbf{H}^+ \mathbf{y}$, where $\mathbf{H}^+$ is the pseudoinverse of $\mathbf{H}$. In the overdetermined case, the estimate $\hat{\underline{\boldsymbol{\theta}}}$ is an unbiased estimate of $\underline{\boldsymbol{\theta}}$. However, in the underdetermined case, $\hat{\underline{\boldsymbol{\theta}}}$ is an unbiased estimate of a rank-N projection of $\underline{\boldsymbol{\theta}}$ onto the subspace spanned by the N columns of $\mathbf{V}$. This projection also happens to be the minimum norm solution to the equations $\mathbf{y} = \mathbf{H}\underline{\boldsymbol{\theta}}$.

When we compute the least squares estimate of $\mathbf{x}$, we are solving the signal extraction or filtering problem. The estimate $\hat{\mathbf{x}}$ of the signal $\mathbf{x}$ is $\hat{\mathbf{x}} = \mathbf{H}\hat{\underline{\boldsymbol{\theta}}}$. In the underdetermined case, $\hat{\mathbf{x}} = \mathbf{y}$. That is, the observation is reproduced exactly. For the overdetermined case, $\hat{\mathbf{x}}$ is the rank-p projection of $\mathbf{y}$ onto the subspace $\langle \mathbf{H} \rangle$. This solution minimizes $(\mathbf{y}-\hat{\mathbf{x}})^T(\mathbf{y}-\hat{\mathbf{x}})$. In this paper, we explore ways of replacing $\hat{\underline{\boldsymbol{\theta}}}$ and $\hat{\mathbf{x}}$ with rank-shaped approximations. See Fig. 1 for a pictorial interpretation.

If we scrutinize the solution to the inversion problem by writing $\mathbf{H}^+$ in its SVD form $\mathbf{V}\underline{\boldsymbol{\Sigma}}^{-1}\mathbf{U}^T$, we get the decomposition

$$\hat{\underline{\boldsymbol{\theta}}} = \mathbf{H}^+ \mathbf{y} = \mathbf{V}\underline{\boldsymbol{\Sigma}}^{-1}\mathbf{U}^T \mathbf{y} = \sum_i \mathbf{v}_i \frac{1}{\sigma_i} \mathbf{u}_i^T \mathbf{y} \tag{2.1}$$

where $\mathbf{U}$ and $\mathbf{V}$ consist of columns $\mathbf{u}_i$ and $\mathbf{v}_i$, respectively. We see that the solution may be noise sensitive because of small singular values σ_i in the SVD decomposition of $\mathbf{H}$ [4]. Small singular values imply that $\mathbf{H}$ is ill-conditioned, a common phenomenon in inverse problems such as numerical deconvolution [5].

What if we replace $\mathbf{H}^+$ by a "rank-shaped" version of $\mathbf{H}^+$, which we denote $\mathbf{H}_r^+$? What effect will this have on the parameter and signal estimates? In particular, can we reduce some measure of the error between $\hat{\underline{\boldsymbol{\theta}}}$ and $\underline{\boldsymbol{\theta}}$ by appropriately choosing $\mathbf{H}_r^+$? Reducing the rank of $\mathbf{H}$ is sometimes referred to as truncating the SVD and shaping the rank is called regularizing $\mathbf{H}^+$. Other techniques have been proposed that regularize the pseudoinverse so that the solution is smooth [1], [7].

Before proceeding with our study of suitable approximations $\mathbf{H}_r^+$, let us establish our conventions:

$\mathbf{H}^+ = \mathbf{V}\underline{\boldsymbol{\Sigma}}^{-1}\mathbf{U}^T$ is the pseudoinverse of $\mathbf{H}$;

$\mathbf{H}_r = \mathbf{U}\underline{\boldsymbol{\Sigma}}_r\mathbf{V}^T$ is the "rank-shaped" version of $\mathbf{H}$;

$\underline{\boldsymbol{\Sigma}}_r = \underline{\boldsymbol{\Gamma}}^{-1}\underline{\boldsymbol{\Sigma}}$ is a diagonal matrix of weighted singular values;

$\underline{\boldsymbol{\Gamma}} = \operatorname{diag}\{\gamma_1, \cdots, \gamma_m\}$ is a diagonal matrix of non-negative weights, $0 < \gamma_i \leq 1$.

$\mathbf{H}_r^+ = \mathbf{V}\underline{\boldsymbol{\Sigma}}_r^{-1}\mathbf{U}^T = \mathbf{V}\underline{\boldsymbol{\Gamma}}\underline{\boldsymbol{\Sigma}}^{-1}\mathbf{U}^T$ is the "rank-shaped" pseudoinverse of $\mathbf{H}$;

$N[\mathbf{m}, \mathbf{R}]$ is a normal distribution with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{R}$.

A. Rank-Shaped Inversion

We begin with the rank-reduced estimate of the parameter vector $\underline{\boldsymbol{\theta}}$

$$\hat{\underline{\boldsymbol{\theta}}}_r = \mathbf{H}_r^+ \mathbf{y} = \mathbf{V}\underline{\boldsymbol{\Gamma}}\underline{\boldsymbol{\Sigma}}^{-1}\mathbf{U}^T \mathbf{y} = \sum_i \mathbf{v}_i \frac{\gamma_i}{\sigma_i} \mathbf{u}_i^T \mathbf{y} \tag{2.2}$$

which is distributed as $N[\mathbf{H}_r^+ \mathbf{x}, \sigma^2 \mathbf{H}_r^+ \mathbf{H}_r^{+T}]$. We must ask ourselves how good this estimator is, for it is no longer unbiased. We shall define the error in $\hat{\underline{\boldsymbol{\theta}}}_r$ as

$$\hat{\underline{\boldsymbol{\xi}}}_r = \tilde{\underline{\boldsymbol{\theta}}} - \hat{\underline{\boldsymbol{\theta}}}_r; \qquad \tilde{\underline{\boldsymbol{\theta}}} = \mathbf{H}^+\mathbf{H}\underline{\boldsymbol{\theta}}. \tag{2.3}$$

Note that $\tilde{\underline{\boldsymbol{\theta}}}$ is just $\underline{\boldsymbol{\theta}}$ in the overdetermined case since $\mathbf{H}^+\mathbf{H} = \mathbf{I}$. In other words, $\hat{\underline{\boldsymbol{\xi}}}_r$ is just $\underline{\boldsymbol{\theta}} - \hat{\underline{\boldsymbol{\theta}}}_r$. In the underdetermined case, $\mathbf{H}^+\mathbf{H} \neq \mathbf{I}$. In fact, $\tilde{\underline{\boldsymbol{\theta}}}$ is a projection of $\underline{\boldsymbol{\theta}}$, which is the minimum norm solution for the underdetermined problem when there is no noise. For both cases, the error $\hat{\underline{\boldsymbol{\xi}}}_r$ is distributed as

$$\hat{\underline{\boldsymbol{\xi}}}_r : N[\mathbf{c}_r, \sigma^2 \mathbf{H}_r^+ \mathbf{H}_r^{+T}] \tag{2.4}$$

where $\mathbf{c}_r$ is the unknown bias

$$\begin{aligned} \mathbf{c}_r &= (\mathbf{H}^+\mathbf{H} - \mathbf{H}_r^+\mathbf{H})\underline{\boldsymbol{\theta}} \\ &= \mathbf{V}(\underline{\boldsymbol{\Gamma}} - \mathbf{I})\underline{\boldsymbol{\Sigma}}^{-1}\mathbf{U}^T\mathbf{x}. \end{aligned} \tag{2.5}$$

The MSE is

$$\begin{aligned} \xi_r^2 &= E(\hat{\underline{\boldsymbol{\xi}}}_r^T \hat{\underline{\boldsymbol{\xi}}}_r) = \mathbf{c}_r^T \mathbf{c}_r + \sigma^2 \operatorname{tr} \mathbf{H}_r^+ \mathbf{H}_r^{+T} \\ &= \mathbf{c}_r^T \mathbf{c}_r + \sigma^2 \operatorname{tr}(\Gamma \underline{\boldsymbol{\Sigma}}^{-1})^2 \\ &= \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \left[[(\gamma_i - 1)\mathbf{u}_i^T \mathbf{x}]^2 + \sigma^2 \left(\frac{\gamma_i}{\sigma_i} \right)^2 \right]. \end{aligned} \tag{2.6}$$

The last line consists of two terms, both quadratic in γ_i. The first term is a bias-squared term that is minimized when $\underline{\boldsymbol{\Gamma}} = \mathbf{I}$ or $\gamma_i = 1 \ \forall i$. The second term is a variance term that is minimized when $\underline{\boldsymbol{\Gamma}} = \mathbf{0}$ or $\gamma_i = 0 \ \forall i$. *Thus, the minimization of ξ_r^2 with respect to the weights γ_i already presents us with a classic tradeoff between squared bias and variance.* Since the sum of two convex upward parabolas must have a global minimum that lies between the global minima of the individually summed parabolas, we see that γ_is outside the range [0, 1] will never minimize the MSE between $\tilde{\underline{\boldsymbol{\theta}}}$ and $\hat{\underline{\boldsymbol{\theta}}}_r$.

Consequently, any procedure designed to minimize ξ_r^2 with respect to γ_i should constrain estimates of γ_i to be in the range [0, 1].

B. Rank-Shaped Filtering

We should ask ourselves another question: if we use a rank-shaped parameter estimate $\hat{\underline{\theta}}_r$, how well do we reproduce the signal component $\mathbf{x}$ of the observation vector $\mathbf{y}$? In other words, with $\hat{\mathbf{x}}_r = \mathbf{H}\hat{\underline{\theta}}_r$, what can we say about the error $\hat{\mathbf{e}}_r = \mathbf{x} - \hat{\mathbf{x}}_r$ and its mean-squared value $e_r^2 = E(\hat{\mathbf{e}}_r^T\hat{\mathbf{e}}_r)$? The error $\hat{\mathbf{e}}_r$ is distributed normally as

$$\hat{\mathbf{e}}_r = \mathbf{x} - \hat{\mathbf{x}}_r = \mathbf{H}(\underline{\theta} - \hat{\underline{\theta}}_r)\colon N[\mathbf{b}_r, \sigma^2\mathbf{H}\mathbf{H}_r^+\mathbf{H}_r^{+T}\mathbf{H}^T] \quad (2.7)$$

where the unknown bias $\mathbf{b}_r$ is

$$\mathbf{b}_r = (\mathbf{I} - \mathbf{H}\mathbf{H}_r^+)\mathbf{x} = (\mathbf{I} - \mathbf{U}\underline{\Gamma}\mathbf{U}^T)\mathbf{x}. \quad (2.8)$$

The MSE is

$$\begin{aligned} e_r^2 &= E(\hat{\mathbf{e}}_r^T\hat{\mathbf{e}}_r) \\ &= \mathbf{b}_r^T\mathbf{b}_r + \sigma^2\mathrm{tr}\mathbf{H}\mathbf{H}_r^+\mathbf{H}_r^{+T}\mathbf{H}^T = \mathbf{b}_r^T\mathbf{b}_r + \sigma^2\mathrm{tr}\underline{\Gamma}^2 \\ &= \sum_{i=1}^{m}\left[(x_i^2 - 2\gamma_i(\mathbf{u}_i^T\mathbf{x})^2 + (\gamma_i\mathbf{u}_i^T\mathbf{x})^2) + (\sigma^2\gamma_i^2)\right]. \end{aligned} \quad (2.9)$$

The last line consists of two terms, both quadratic in $\underline{\Gamma}$. The first term is a squared bias term that is minimized when $\underline{\Gamma} = \mathbf{I}$ or $\gamma_i = 1\ \forall i$. The second term is a variance term that is minimized when $\Gamma = \mathbf{0}$ or $\gamma_i = 0\ \forall i$. *Again, we have a classic bias-squared versus variance trade where γ_is outside the range* [0, 1] *will never minimize the MSE between* $\mathbf{x}$ *and* $\hat{\mathbf{x}}_r$. Consequently, when we minimize e_r^2 with respect to γ_i, we should constrain our estimates of γ_i to be in the range [0, 1].

III. Clairvoyant Estimates of the Weighting Coefficients

The results of the previous section bring insight into the dependence of MSE on the weighting coefficients γ_i. The solutions for coefficients that minimize MSE are only idealized results because they depend on clairvoyant knowledge of $\mathbf{x}$ or $\underline{\theta}$, which of course we do not have. Nonetheless, by studying these idealized solutions we gain insight into suitable data-adaptive solutions.

A. Least Squares

In the inversion problem, the dependence of MSE ξ_r^2 on $\mathbf{x}$ is given in (2.6). Differentiating ξ_r^2 with respect to γ_i and equating the result to 0 yields the clairvoyant solution

$$\frac{\partial\xi_r^2}{\partial\gamma_i} = 0 = \frac{2}{\sigma_i^2}[(\gamma_i - 1)(\mathbf{u}_i^T\mathbf{x})^2 + \gamma_i\sigma^2] \quad (3.1)$$

$$\gamma_i = \frac{(\mathbf{u}_i^T\mathbf{x})^2}{(\mathbf{u}_i^T\mathbf{x})^2 + \sigma^2}; \qquad i = 1, \cdots, m.$$

Likewise, for the filtering problem, minimizing e_r^2 in (2.9) leaves us with the identical solution for the clairvoyant γ_is

$$\frac{\partial e_r^2}{\partial\gamma_i} = 0 = 2\left[-(\mathbf{u}_i^T\mathbf{x})^2 + \gamma_i(\mathbf{u}_i^T\mathbf{x})^2 + \gamma_i\sigma^2\right] \quad (3.2)$$

$$\gamma_i = \frac{(\mathbf{u}_i^T\mathbf{x})^2}{(\mathbf{u}_i^T\mathbf{x})^2 + \sigma^2}; \qquad i = 1, \cdots, m.$$

These coefficients take on values in the range [0, 1]. What is more important, however, is the fact that *each clairvoyant γ_i is just the ratio of the power in the ith mode of the signal* $\mathbf{x}$ *to the power in the ith mode of the observation* $\mathbf{y}$. In fact, γ_i can also be written as the ratio

$$0 \le \gamma_i = \frac{\beta_i^2}{\beta_i^2 + 1} \le 1 \quad (3.3)$$

where the β_i^2 are SNR's in the respective modes:

$$\beta_i^2 = (\mathbf{u}_i^T\mathbf{x})^2/\sigma^2 = (\sigma_i\mathbf{v}_i^T\underline{\theta})^2/\sigma^2 = \frac{\gamma_i}{1 - \gamma_i}. \quad (3.4)$$

B. The Wiener Weighting Coefficients

Consider the Wiener solution to the inversion and filtering problems. Assuming means of zero, the Wiener solutions to these problems are

$$\begin{aligned} \hat{\underline{\theta}} &= \mathbf{R}_{\theta y}\mathbf{R}_{yy}^{-1}\mathbf{y} \\ \hat{\mathbf{x}} &= \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{y} \end{aligned} \quad (3.5)$$

where $\mathbf{R}_{uv} = E\mathbf{u}\mathbf{v}^T$. The Wiener solution requires knowledge of the cross-covariance structure that relates $\mathbf{y}$ to $\underline{\theta}$ or $\mathbf{y}$ to $\mathbf{x}$, but we have made no assumption thus far about the statistical nature of $\underline{\theta}$ and $\mathbf{x}$. Let's assume that $\underline{\theta}$ has covariance $\sigma_S^2\mathbf{I}$

$$\begin{aligned} \mathbf{R}_{\theta\theta} &= \sigma_S^2\mathbf{I} \\ E\theta_i\theta_j &= \sigma_S^2\delta_{ij}. \end{aligned} \quad (3.6)$$

With this assumption, we compute the Wiener solution by solving for $\mathbf{R}_{yy}$, $\mathbf{R}_{\theta y}$, and $\mathbf{R}_{xy}$:

$$\begin{aligned} \mathbf{R}_{yy} &= \mathbf{H}\mathbf{R}_{\theta\theta}\mathbf{H}^T + \mathbf{R}_{nn}; \qquad \mathbf{R}_{\theta y} = \mathbf{R}_{\theta\theta}\mathbf{H}^T; \\ \mathbf{R}_{xy} &= \mathbf{H}\mathbf{R}_{\theta\theta}\mathbf{H}^T. \end{aligned} \quad (3.7)$$

From these results we estimate $\underline{\theta}$ and $\mathbf{x}$ as follows:

$$\begin{aligned} \hat{\underline{\theta}} &= \mathbf{R}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{R}_{\theta\theta}\mathbf{H}^T + \mathbf{R}_{nn})^{-1}\mathbf{y} \\ &= (\mathbf{R}_{\theta\theta}^{-1} + \mathbf{H}^T\mathbf{R}_{nn}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}_{nn}^{-1}\mathbf{y} \\ \hat{\mathbf{x}} &= \mathbf{H}\hat{\underline{\theta}}. \end{aligned} \quad (3.8)$$

For both the underdetermined and the overdetermined problems, the solution for $\underline{\theta}$ may be written

$$\hat{\underline{\theta}} = E[\underline{\theta} \mid \mathbf{y}] = \mathbf{V}\underline{\Sigma}^{-1}\Gamma_W\mathbf{U}^T\mathbf{y} \quad (3.9)$$

where $\mathbf{\Gamma}_W$ is the diagonal matrix

$$\begin{aligned} \mathbf{\Gamma}_W &= \sigma_S^2\underline{\Sigma}^2(\sigma_S^2\underline{\Sigma}^2 + \sigma^2\mathbf{I})^{-1} \\ &= \mathrm{diag}[[\gamma_i]_W]; \qquad [\gamma_i]_W = \sigma_S^2\sigma_i^2/(\sigma_S^2\sigma_i^2 + \sigma^2). \end{aligned} \quad (3.10)$$

That is

$$\hat{\underline{\theta}} = \sum_i \mathbf{v}_i\frac{[\gamma_i]_W}{\sigma_i}\mathbf{u}_i^T\mathbf{y}. \quad (3.11)$$

The Wiener solutions for $\hat{\underline{\theta}}$ and $\hat{\mathbf{x}}$ are of the same form as the rank-shaped least squares solutions of (2.2)! The diagonal weighting matrix $\mathbf{\Gamma}_W$ smoothly shapes the rank of $\mathbf{H}^+$ in order to get the minimum MSE estimates of $\underline{\theta}$ and $\mathbf{x}$. Moreover, the $[\gamma_i]_W$ coefficients have values in the range [0, 1] and are of the same form as the clairvoyant least squares γ_i coefficients computed in (3.3)

$$[\gamma_i]_W = \frac{\sigma_S^2\sigma_i^2}{\sigma_S^2\sigma_i^2+\sigma^2} = \frac{[\beta_i^2]_W}{[\beta_i^2]_W+1} \tag{3.12}$$

with the following definition of SNR:

$$[\beta_i^2]_W = \sigma_s^2\sigma_i^2/\sigma^2 = E[(\mathbf{u}_i^T\mathbf{x})^2/\sigma^2] = E[(\sigma_i\mathbf{v}_i^T\underline{\boldsymbol{\theta}})^2/\sigma^2] = \frac{[\gamma_i]_W}{1-[\gamma_i]_W}. \tag{3.13}$$

There is *no essential difference between the clairvoyant least squares solution of* (3.3) *and* (3.4) *and the Wiener solution of* (3.12) *and* (3.13): *they both use rank shaping with shape parameter* $0 \le \gamma_i \le 1$.

The least squares coefficients γ_i are a clairvoyant solution to the least squares problem. For a fixed $\underline{\boldsymbol{\theta}}$ and multiple noise realizations, this solution will provide, on average, the smallest MSE estimates of $\underline{\boldsymbol{\theta}}$ and $\mathbf{x}$. The Wiener coefficients $[\gamma_i]_W$ are computed when the covariance matrix for $\underline{\boldsymbol{\theta}}$ is known but $\underline{\boldsymbol{\theta}}$ itself is not. The Wiener solution gives us the minimum MSE estimates of the signal or parameter vector for multiple signal-plus-noise realizations. Equation (3.13) shows that the realizable Wiener solution uses the average power in the ith mode of the parameter, whereas the unrealizable clairvoyant least squares solution of (3.3) uses the exact power in the ith mode of the parameter. This observation is insightful but not yet useful.

IV. Naive Estimates of the Weighting Coefficients

The clairvoyant solutions for minimizing weighting coefficients depend upon exact knowledge of the mean-squared errors ξ_r^2 and e_r^2, which in turn depend on exact knowledge of the signal $\mathbf{x}$. Any practical, data-dependent solution for weighting coefficients must rely on estimates of what is unknown, not on exact knowledge. The primitive estimates we study in this section use data-dependent estimates of the mean-squared errors ξ_r^2 and e_r^2 to derive data-dependent estimates of the weighting coefficients.

A. Unbiased, Constrained, and Abrupt Estimates

Recall that the MSE ξ_r^2 of (2.6) is decomposed into bias-squared plus variance, with bias unknown. An unbiased estimate of the bias $\mathbf{c}_r$ that is valid for both underdetermined and overdetermined cases is

$$\hat{\mathbf{c}}_r = (\mathbf{H}^+ - \mathbf{H}_r^+)\mathbf{y}\colon N[\mathbf{c}_r, \sigma^2(\mathbf{H}^+ - \mathbf{H}_r^+)(\mathbf{H}^+ - \mathbf{H}_r^+)^T]. \tag{4.1}$$

The variance of the estimator $\hat{\mathbf{c}}_r$ is

$$\begin{aligned} E(\hat{\mathbf{c}}_r - \mathbf{c}_r)^T(\hat{\mathbf{c}}_r - \mathbf{c}_r) &= E\hat{\mathbf{c}}_r^T\hat{\mathbf{c}}_r - \mathbf{c}_r^T\mathbf{c}_r \\ &= \sigma^2\mathrm{tr}(\mathbf{H}^+ - \mathbf{H}_r^+)(\mathbf{H}^+ - \mathbf{H}_r^+)^T \\ &= \sigma^2\mathrm{tr}\underline{\boldsymbol{\Sigma}}^{-2}(\underline{\boldsymbol{\Gamma}} - \mathbf{I})^2. \end{aligned} \tag{4.2}$$

This produces the fundamental identity

$$E\hat{\mathbf{c}}_r^T\hat{\mathbf{c}}_r = \mathbf{c}_r^T\mathbf{c}_r + \sigma^2\mathrm{tr}\underline{\boldsymbol{\Sigma}}^{-2}(\underline{\boldsymbol{\Gamma}} - \mathbf{I})^2. \tag{4.3}$$

This identity shows that $\hat{\mathbf{c}}_r^T\hat{\mathbf{c}}_r$ is a *biased estimator* of the squared bias $\mathbf{c}_r^T\mathbf{c}_r$ even though $\hat{\mathbf{c}}_r$ is an *unbiased* estimate of the bias $\mathbf{c}_r$!

With this result, we can form the following unbiased estimate of the MSE ξ_r^2:

$$\begin{aligned} \hat{\xi}_r^2 &= \hat{\mathbf{c}}_r^T\hat{\mathbf{c}}_r - \sigma^2\mathrm{tr}\underline{\boldsymbol{\Sigma}}^{-2}(\underline{\boldsymbol{\Gamma}} - \mathbf{I})^2 + \sigma^2\mathrm{tr}(\underline{\boldsymbol{\Gamma}}\underline{\boldsymbol{\Sigma}}^{-1})^2 \\ &= \hat{\mathbf{c}}_r^T\hat{\mathbf{c}}_r + \sigma^2\mathrm{tr}(2\underline{\boldsymbol{\Gamma}} - \mathbf{I})\underline{\boldsymbol{\Sigma}}^{-2}. \end{aligned} \tag{4.4}$$

This expression may be expanded as follows:

$$\begin{aligned} \hat{\xi}_r^2 &= \|(\mathbf{H}^+ - \mathbf{H}_r^+)\mathbf{y}\|^2 + \sigma^2\mathrm{tr}(2\underline{\boldsymbol{\Gamma}} - \mathbf{I})\underline{\boldsymbol{\Sigma}}^{-2} \\ &= \sum_{i=1}^{m}\frac{1}{\sigma_i^2}[[(\gamma_i - 1)\mathbf{u}_i^T\mathbf{y}]^2 + \sigma^2(2\gamma_i - 1)]. \end{aligned} \tag{4.5}$$

Let's follow the philosophy of the clairvoyant estimator to minimize *estimated* MSE with respect to the weighting coefficients γ_i under the constraint that $0 \le \gamma_i \le 1$.

The constrained minimization of $\hat{\xi}_r^2$ is obtained by computing partial derivatives and equating them to zero. Since $\hat{\xi}_r^2$ is quadratic in each γ_i and there are no cross terms between the γ_is, the constraint can be applied after using an unconstrained minimization of $\hat{\xi}_r^2$. The unconstrained minimum is obtained as follows:

$$\frac{\partial\hat{\xi}_r^2}{\partial\gamma_i} = 0 = \frac{2}{\sigma_i^2}[(\hat{\gamma}_i - 1)(\mathbf{u}_i^T\mathbf{y})^2 + \sigma^2] \tag{4.6}$$

$$\hat{\gamma}_i = \frac{\hat{\beta}_i^2 - 1}{\hat{\beta}_i^2}; \qquad \hat{\beta}_i^2 = (\mathbf{u}_i^T\mathbf{y})^2/\sigma^2; \qquad i = 1, \cdots, m.$$

Let's compare the estimated $\hat{\gamma}_i$ of (4.6) with the clairvoyant solution, $\gamma_i = \beta_i^2/(\beta_i^2 + 1)$. Is the comparison plausible? The estimator $\hat{\beta}_i^2$ is a biased estimator of the SNR β_i^2, as the following argument shows. $\hat{\beta}_i = \mathbf{u}_i^T\mathbf{y}/\sigma$ is distributed as $N[\beta_i, 1]$, meaning that $E\hat{\beta}_i^2 = \beta_i^2 + 1$. Therefore, $\hat{\gamma}_i$ may be written as

$$\hat{\gamma}_i = \frac{[\hat{\beta}_i^2]_{\mathrm{UB}}}{[\hat{\beta}_i^2]_{\mathrm{UB}} + 1} \tag{4.7}$$

where $[\hat{\beta}_i^2]_{\mathrm{UB}} = \hat{\beta}_i^2 - 1$ is an unbiased estimate of β_i^2. *This shows that the minimization of the estimated MSE produces the same answer we would get if we just replaced the per-mode SNR's in the clairvoyant solution for* γ_i *with unbiased estimates of the per-mode SNR's.* This seems plausible, but, as we shall show, it is not reasonable for it produces poor performance.

We may follow these arguments for the minimization of the estimated MSE $\hat{e}_r^2$ as well. Recall that the MSE e_r^2 of (2.8) is decomposed into squared bias plus variance, with bias unknown.

An unbiased estimator of $\mathbf{b}_r$ is

$$\hat{\mathbf{b}}_r = (\mathbf{I} - \mathbf{H}\mathbf{H}_r^+)\mathbf{y}\colon N[\mathbf{b}_r, \sigma^2(\mathbf{I} - \mathbf{H}\mathbf{H}_r^+)(\mathbf{I} - \mathbf{H}\mathbf{H}_r^+)^T]. \tag{4.8}$$

The variance of the estimator $\hat{\mathbf{b}}_r$ is

$$\begin{aligned} E(\hat{\mathbf{b}}_r - \mathbf{b}_r)^T(\hat{\mathbf{b}}_r - \mathbf{b}_r) &= E\hat{\mathbf{b}}_r^T\hat{\mathbf{b}}_r - \mathbf{b}_r^T\mathbf{b}_r \\ &= \sigma^2\mathrm{tr}(\mathbf{I} - \mathbf{H}\mathbf{H}_r^+)(\mathbf{I} - \mathbf{H}\mathbf{H}_r^+)^T \\ &= \sigma^2\mathrm{tr}(\mathbf{I} - \mathbf{U}\underline{\boldsymbol{\Gamma}}\mathbf{U}^T)(\mathbf{I} - \mathbf{U}\underline{\boldsymbol{\Gamma}}\mathbf{U}^T)^T. \end{aligned} \tag{4.9}$$

This computation shows that $\hat{\mathbf{b}}_r^T\hat{\mathbf{b}}_r$ is a *biased estimate* of $\mathbf{b}_r^T\mathbf{b}_r$, even though $\hat{\mathbf{b}}_r$ is an *unbiased estimate* of $\mathbf{b}_r$:

$$E\hat{\mathbf{b}}_r^T\hat{\mathbf{b}}_r = \mathbf{b}_r^T\mathbf{b}_r + \sigma^2\text{tr}(\mathbf{I} - \mathbf{U}\Gamma\mathbf{U}^T)^2 = \mathbf{b}_r^T\mathbf{b}_r + \sigma^2\text{tr}(\mathbf{I} - 2\underline{\Gamma} + \underline{\Gamma}^2). \quad (4.10)$$

Now, we can form the following unbiased estimate of the MSE e_r^2:

$$\hat{e}_r^2 = \hat{\mathbf{b}}_r^T\hat{\mathbf{b}}_r - \sigma^2\text{tr}(\mathbf{I} - 2\underline{\Gamma} + \underline{\Gamma}^2) + \sigma^2\text{tr}\Gamma^2 = \hat{\mathbf{b}}_r^T\hat{\mathbf{b}}_r + \sigma^2\text{tr}(2\underline{\Gamma} - \mathbf{I}). \quad (4.11)$$

Expanding this expression gives

$$\hat{e}_r^2 = \|(\mathbf{I} - \mathbf{H}\mathbf{H}_r^+)\mathbf{y}\|^2 + \sigma^2\text{tr}(2\underline{\Gamma} - \mathbf{I}) = \sum_{i=1}^{m}[y_i^2 - 2\gamma_i(\mathbf{u}_i^T\mathbf{y})^2 + (\gamma_i\mathbf{u}_i^T\mathbf{y})^2 + \sigma^2(2\gamma_i - 1)]. \quad (4.12)$$

Using the same procedure as before, we minimize $\hat{e}_r^2$ by computing partial derivatives and equating them to zero. Since $\hat{e}_r^2$ is quadratic in each γ_i and there are no cross terms between the γ_is, the constraint can be applied after using an unconstrained minimization of $\hat{e}_r^2$. The unconstrained minimum is achieved as follows:

$$\frac{\partial \hat{e}_r^2}{\partial \gamma_i} = 0 = 2[(\hat{\gamma}_i - 1)(\mathbf{u}_i^T\mathbf{y})^2 + \sigma^2]$$

$$\hat{\gamma}_i = \frac{\hat{\beta}_i^2 - 1}{\hat{\beta}_i^2}; \qquad i = 1, \cdots, m. \quad (4.13)$$

This solution is identical to the solution to minimize the estimator error in the parameter estimate $\hat{\underline{\theta}}_r$. This analysis tells us to use exactly the same rank-shaping principles when we minimize the MSE of our solution regardless of whether we are solving the inversion or filtering problem.

These results extend in the following way to the more common approach of abrupt rank *reduction*, wherein the weighting coefficients γ_i have values of either zero or one. Each of the estimated error expressions $\hat{\xi}_r^2$ and $\hat{e}_r^2$ define a multidimensional surface that is quadratic in the γ_is. Therefore, the best abrupt rank reduction is obtained by thresholding the $\hat{\gamma}_i$s

$$[\hat{\gamma}_i]_A = T(\hat{\gamma}_i) \quad (4.14)$$

where

$$T(x) = \begin{cases} 1, & \text{if } x \geq 1/2 \\ 0, & \text{otherwise.} \end{cases} \quad (4.15)$$

Thresholding the $\hat{\gamma}_i$s yields the point on a corner of the m-dimensional hypercube where the error estimates are smallest. This result improves on a procedure for abrupt rank reduction proposed in [8]–[10].

We complete our derivation of naive estimators of the weighting coefficients by enforcing the nonnegative constraints

$$\hat{\gamma}_i = \max\left[0, \frac{\hat{\beta}_i^2 - 1}{\hat{\beta}_i^2}\right]. \quad (4.16)$$

The corresponding rank-shaped estimators for $\underline{\theta}$ and $\mathbf{x}$ are

$$\hat{\underline{\theta}}_r = \mathbf{V}\hat{\Gamma}\underline{\Sigma}^{-1}\mathbf{U}^T\mathbf{y} = \sum_i \mathbf{v}_i\frac{\hat{\gamma}_i}{\sigma_i}\mathbf{u}_i^T\mathbf{y};$$

$$\hat{\mathbf{x}}_r = \mathbf{U}\hat{\Gamma}\mathbf{U}^T\mathbf{y} = \sum_i \mathbf{u}_i\hat{\gamma}_i\mathbf{u}_i^T\mathbf{y}. \quad (4.17)$$

For abrupt rank reduction, these estimators are

$$\hat{\underline{\theta}}_r = \sum_{i\in\mathbf{I}} \mathbf{v}_i\frac{1}{\sigma_i}\mathbf{u}_i^T\mathbf{y}; \qquad \hat{\mathbf{x}}_r = \sum_{i\in\mathbf{I}} \mathbf{u}_i\mathbf{u}_i^T\mathbf{y} \quad (4.18)$$

where $\mathbf{I}$ is the index set for which $\hat{\gamma}_i > 1/2$. These are purely rank-reduced pseudoinverses and projections. Note that even in the case of abrupt rank reduction, the solutions are highly nonlinear in the data by virtue of the nonlinear dependence of $\hat{\gamma}_i$ on the data $\mathbf{y}$.

B. Companders and Performance

Each of the estimators $\hat{\gamma}_i$ and $[\hat{\gamma}_i]_A$ is a function of the estimated SNR $\hat{\beta}_i^2$, which has values in the range $[0, \infty)$. This infinite range of values is inconvenient for fixed-point computations, and therefore we consider the variable z_i defined as

$$z_i = \frac{\hat{\beta}_i^2}{\hat{\beta}_i^2 + 1}; \qquad \hat{\beta}_i^2 = \frac{z_i}{1 - z_i}. \quad (4.19)$$

Notice that z_i has values in the finite range [0, 1] and that the function that defines z_i is invertible for $\hat{\beta}_i^2$. Consequently, all the estimates of γ_i can be written as functions of z_i instead of as functions of $\hat{\beta}_i^2$. These functions can be thought of as companders that operate on the interval [0, 1], the range of z_i. That is

$$\hat{\gamma}_i = C_0(z_i) = \max\left[0, 2 - \frac{1}{z_i}\right] \quad (4.20)$$

$$[\hat{\gamma}_i]_A = C_1(z_i) = \begin{cases} 1, & \text{if } z_i \geq 2/3 \\ 0, & \text{otherwise.} \end{cases} \quad (4.21)$$

The companders C_0 and C_1 are plotted for comparison in Fig. 2 and compared with the maximum likelihood compander to be derived in Section 5-A.

The conditional mean and MSE of each of these estimators are

$$E[C_j(z_i)] = \int_0^1 C_j(t)f_{z_i}(t)\,dt; \qquad j = 0, 1 \quad (4.22)$$

$$E[(C_j(z_i) - \gamma_i)^2] = \int_0^1 (C_j(t) - \gamma_i)^2 f_{z_i}(t)\,dt; \qquad j = 0, 1. \quad (4.23)$$

Of course, we must compute the density function for z_i for fixed γ_i, because the data-dependent z_i is just a coarse estimate of the unknown clairvoyant gain. This density function is identical to the conditional density function of z_i given γ_i, which is computed as part of the appendix, the result being (6) of the appendix. With this result, the mean and MSE of each estimator $\hat{\gamma}_i$ of γ_i can be computed numerically as a function of γ_i. The conditional mean for each of the estimators is plotted in Fig. 3 versus γ_i. If there existed a conditionally

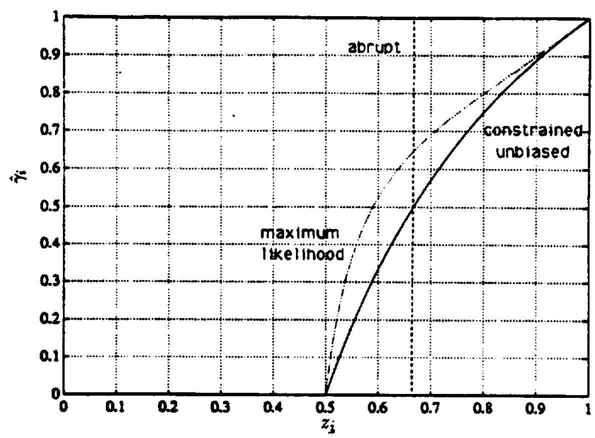


Fig. 2. Constrained unbiased, abrupt, and maximum likelihood estimators of γ_i.

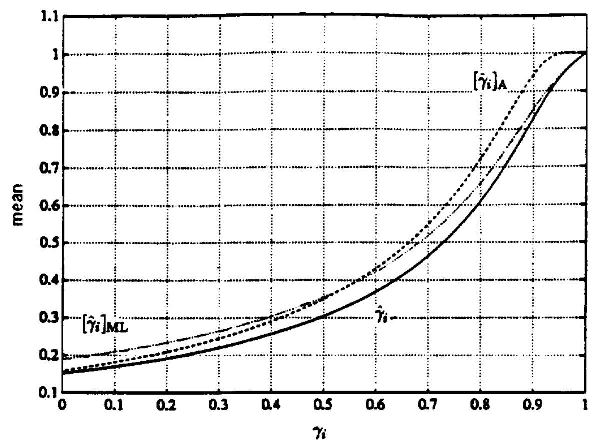


Fig. 3. Mean of the constrained unbiased, abrupt, and maximum likelihood estimators as a function of γ_i.

unbiased estimator, it would show up on the graph as a straight line connecting (0, 0) and (1, 1). We see that both estimators have a bad positive bias when γ_i is small. This means that a mode with a small singular value is likely to be used in the solution with more weight than deserved, degrading the result. All estimators do, however, have relatively small biases for γ_i close to one.

Plotted in Fig. 4 is the MSE for the estimators $\hat{\gamma}_i$ and $[\hat{\gamma}_i]_A$ as a function of γ_i. When little rank shaping is required in a mode—i.e., when the true γ_i is close to one—the estimators perform well. However, the estimators of γ_i are poor for most of the range of γ_i. This poor performance leads to poor performance of the rank-reduced estimators $\hat{\underline{\theta}}_r$ and $\hat{\mathbf{x}}_r$, as reported in [5]. Indeed, the quality of the solution in the inversion problem depends on the lowest per-mode SNR β_i^2. The simulations in [11] and [12] indicate that the overall SNR has to be large enough so that the SNR in any one mode is at least 20 dB. Otherwise, the estimates of γ_i in modes with low SNR are very poor, and the noise in those modes degrades the solution. In summary, neither of the realizable rank-shaping methods that use $\hat{\gamma}_i$ or $[\hat{\gamma}_i]_A$ is satisfactory when the model

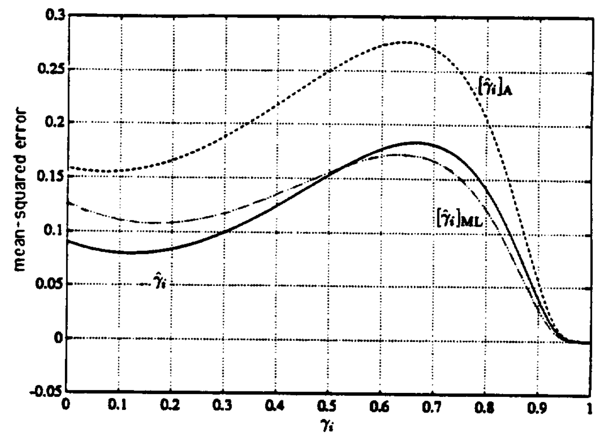


Fig. 4. Mean-squared error of the constrained unbiased, abrupt, and maximum likelihood estimators as a function of γ_i.

$\mathbf{x}$ has low SNR in its subdominant modes. This is a sobering result.

Perhaps we can improve our estimates of γ_i by using the *a priori* information that is available to a Wiener filter. This of course constitutes a fundamental change in direction, for we are proposing to bootstrap ourselves to a useful data adaptive filter by pretending to have a prior distribution on $\underline{\theta}$. Once we are bootstrapped, we will use our results on data that is mismatched to the bootstrapping assumptions. As we shall see, the improvements are remarkable.

V. Sophisticated Estimates of the Weighting Coefficients

We now derive two more estimates of the clairvoyant weighting coefficients γ_i. The first is a maximum likelihood estimate of γ_i, and the second is the conditional mean estimate of γ_i given the measurement $\mathbf{y}$. In order to derive the conditional mean estimator, we assign a prior distribution to the parameter $\underline{\theta}$ as is done in Wiener filtering, determine the resulting prior distribution on $\gamma_i = \sigma_i \mathbf{v}_i^T \underline{\theta}/\sigma$, and use this prior distribution to find the posterior distribution on γ_i given the measurement $\mathbf{y}$. This is *not* the Wiener solution, for it uses the conditional mean of γ_i in a rank-shaped estimator, not the conditional mean of $\underline{\theta}$ or $\mathbf{x}$.

The conditional distribution of $\sigma^{-1}\mathbf{u}^T\mathbf{y} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)^T$ is $\prod_i N(\beta_i, 1)$, with the $\beta_i = \mathbf{u}_i^T\mathbf{x}/\sigma$ the voltage SNR in mode i and $\hat{\beta}_i = \mathbf{u}_i^T\mathbf{y}/\sigma$ the estimated voltage SNR. This makes $\hat{\beta}_i$ sufficient for β_i and $\hat{\beta}_i^2$ sufficient for β_i^2. But $\hat{\beta}_i^2 = z_i/(1 - z_i)$, so z_i is sufficient for β_i^2 and, as $\gamma_i = \beta_i^2/(1 + \beta_i^2)$, z_i is sufficient for γ_i when γ_i is the deterministic but unknown clairvoyant gain. This means that the maximum likelihood estimate of γ_i is a function of the estimated SNR $\hat{\beta}_i^2$ or of the companding variable z_i. By a result in [9, pp. 290–291], $\hat{\beta}_i^2$ and z_i are also Bayes sufficient for γ_i when γ_i is the random parameter induced from $\gamma_i = \beta_i^2/(1+\beta_i^2)$, $\beta_i^2 = (\sigma_i \mathbf{v}_i^T \underline{\theta}/\sigma)^2$, and $\underline{\theta} : N[0, \sigma_s^2 I]$. These results mean that ML and conditional mean estimates of the clairvoyant gain γ_i are companding functions of the bounded variable $0 \le z_i = \hat{\beta}_i^2/(1 + \hat{\beta}_i^2) \le 1$. Think of the companding variablez_i as a coarse estimator of the clairvoyant gain $\gamma_i = \beta_i^2/(1+\beta_i^2)$, which is to be refined by the principles of maximum likelihood and conditional mean.

A. Maximum Likelihood Estimators of the Weighting Coefficients

The function $\gamma_i = \beta_i^2/(1+\beta_i^2)$ satisfies the maximum likelihood invariance requirements [9]. This allows us to compute the maximum likelihood estimate of γ_i using a maximum likelihood estimate of β_i^2 as follows

$$[\hat{\gamma}_i]_{\mathrm{ML}} = \frac{[\hat{\beta}_i^2]_{\mathrm{ML}}}{[\hat{\beta}_i^2]_{\mathrm{ML}} + 1}. \tag{5.1}$$

Recall that $\hat{\beta}_i^2$ is χ^2 distributed. That is, $\hat{\beta}_i^2$: $\chi^2[\beta_i^2, 1]$, where β_i^2 is the noncentrality parameter. Our goal is to find a maximum likelihood estimate of the noncentrality parameter of a χ^2 distribution with one degree of freedom.

First, we will compute the distribution for $\hat{\beta}_i^2$ for $z \geq 0$

$$\begin{aligned} F_{\hat{\beta}_i^2}(\beta^2) &= P[\hat{\beta}_i^2 \leq \beta^2] \\ &= P\left[-\beta \leq \frac{\mathbf{u}_i^T \mathbf{y}}{\sigma} \leq \beta\right]. \end{aligned} \tag{5.2}$$

However, $\mathbf{u}_i^T\mathbf{y}/\sigma$ is distributed as $N[\beta_i, 1]$, so this distribution may be written as

$$F_{\hat{\beta}_i^2}(\beta^2) = \begin{cases} 0, & \text{if } z < 0 \\ \Phi(\beta - \beta_i) - \Phi(-\beta - \beta_i), & \text{otherwise} \end{cases} \tag{5.3}$$

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt : \text{normal integral.}$$

The density function for $\hat{\beta}_i^2$ is the derivative of the distribution function, as shown in (5.4) at the bottom of the page.

Given a non-negative sample β^2 of $\hat{\beta}_i^2$, the maximum likelihood estimate of the noncentrality parameter is

$$\begin{aligned} [\hat{\beta}_i^2]_{\mathrm{ML}} = \arg\max_{\beta_i^2} & \frac{1}{2\sqrt{2\pi\beta^2}} \\ & \times \left[\exp\left(-\frac{1}{2}(\beta - \beta_i)^2\right) + \exp\left(-\frac{1}{2}(\beta + \beta_i)^2\right)\right]. \end{aligned} \tag{5.5}$$

This maximization problem is equivalent to finding the zeros of the derivative of the density function, or

$$\begin{aligned} [\hat{\beta}_i^2]_{\mathrm{ML}} &= \arg\left(\frac{\partial}{\partial \beta_i^2} f_{\hat{\beta}_i^2}(\beta^2) = 0\right) \\ &= \arg\left[\alpha_1 \exp\left(-\frac{\alpha_1^2}{2}\right) - \alpha_2 \exp\left(-\frac{\alpha_2^2}{2}\right) = 0\right]; \\ & \alpha_1 = \beta - \beta_i; \qquad \alpha_2 = \beta + \beta_i. \end{aligned} \tag{5.6}$$

Equation (5.6) can be solved numerically with a zero finding routine for nonlinear functions. Then, once the maximum likelihood estimate for β_i^2 is found, we can compute the maximum likelihood estimate of γ_i via (5.1). The maximum likelihood compander, illustrated in Fig. 2, is very similar to the naive companders for $\hat{\gamma}_i$ and $[\hat{\gamma}_i]_A$. The mean and MSE of $[\hat{\gamma}]_{\mathrm{ML}}$ are plotted in Figs. 3 and 4. It suffers the same defects as the naive companders.

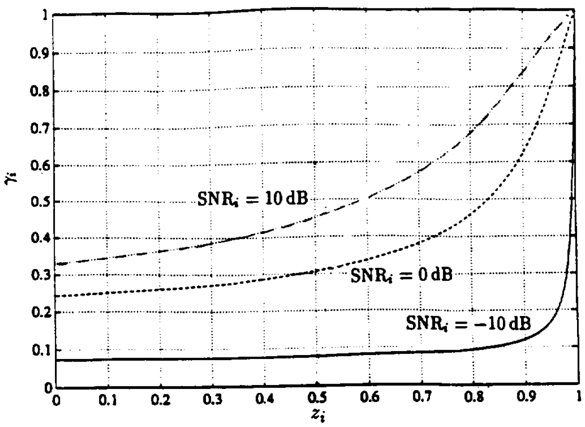


Fig. 5. Conditional mean companders for per-mode SNR's of −10, 0, and 10 dB.

B. Conditional Mean Estimators of the Weighting Coefficients

The conditional mean estimator for γ_i, given the measurement $\mathbf{y}$, is

$$[\hat{\gamma}_i]_{\mathrm{CM}} = E[\gamma_i \mid \mathbf{y}] = E[\gamma_i \mid \hat{\beta}_i^2] = E[\gamma_i \mid z_i] \tag{5.7}$$

where the subscript CM denotes conditional mean. In order to find such an estimator, we need to know the distribution for γ_i conditioned on $\hat{\beta}_i^2$ or z_i. This conditional mean estimator is also the minimum MSE estimator of γ_i and the Bayes estimator of γ_i under quadratic loss.

In the appendix, we derive the conditional density for γ_i given z_i. The density is completely parameterized by the Wiener coefficient $[\gamma_i]_W$. The conditional mean estimator is a function of z_i and can be viewed as a compander that maps z_i to an estimate of the clairvoyant gain γ_i. We must approximate these companders numerically, and we must build a different compander for each mode because each mode has a different SNR. It is interesting to compare these conditional mean companders to the companders that map z_i into $\hat{\gamma}_i$, $[\hat{\gamma}_i]_A$, and $[\hat{\gamma}_i]_{\mathrm{ML}}$. In Fig. 5, we have plotted the conditional mean companders corresponding to per-mode SNR's of −10, 0, and 10 dB. These SNR's reflect a range of singular values of only one order of magnitude or a condition number of only 10^1. The maximum likelihood compander is also plotted in Fig. 5. The conditional mean compander is very different from the previous companders, as the following discussion shows.

Each of the conditional mean companders produces an output γ_i close to one for an input z_i close to one. This is plausible because a large value of z_i is obtained from an observation that has a lot of power in the ith mode. Since the

$$f_{\hat{\beta}_i^2}(\beta^2) = \begin{cases} 0, & \text{if } \beta^2 < 0 \\ \frac{1}{2\sqrt{2\pi\beta^2}}\left[\exp\left(-\frac{1}{2}(\beta - \beta_i)^2\right) + \exp\left(-\frac{1}{2}(\beta + \beta_i)^2\right)\right], & \text{otherwise.} \end{cases} \tag{5.4}$$

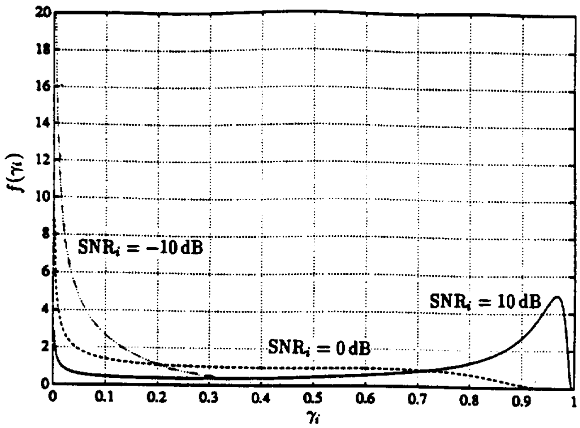


Fig. 6. Probability density function for γ_i at per-mode SNR's of −10, 0, and 10 dB.

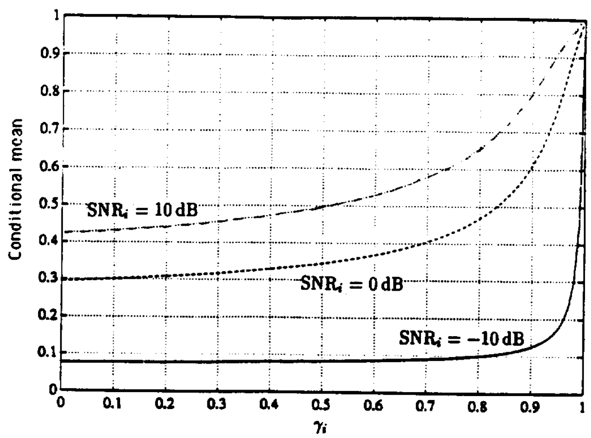


Fig. 7. Conditional mean for the Bayes estimates of γ_i for per-mode SNR's of −10, 0, and 10 dB.

average noise power in a mode is just σ^2, most of the power in the mode must be signal power. Hence, the compander delivers an estimate of γ_i close to one.

Notice also how the −10 dB compander will produce values of $[\hat{\gamma}_i]_{\text{CM}}$ close to $[\gamma_i]_W = .09$ for most values of z_i. Only if there is strong evidence to the contrary—i.e., $z_i > .9$—will the compander produce a much different estimate of γ_i. In general, conditional mean companders for low per-mode SNR's exhibit this characteristic. *That is, an output value close to the Wiener* $[\gamma_i]_W$ *is favored unless the input* z_i *strongly indicates otherwise.* This allows some adaptability to observations $\mathbf{y}$ that are produced by atypical realizations of $\underline{\theta}$. For example, if by chance the realization of $\underline{\theta}$ correlates well with $\mathbf{v}_i^T$, then the signal power will be concentrated in the ith mode and the conditional mean estimator will adjust accordingly.

In (14) of the appendix, we have computed the density for γ_i when $\underline{\theta}$ is distributed as $N[0, \sigma_\theta^2 \mathbf{I}]$. By plotting this density, we can gain additional insight as to why the previously derived constrained unbiased estimators, abrupt estimators, and maximum likelihood estimators fared so poorly. The density is plotted in Fig. 6 for three per-mode SNR's of −10,

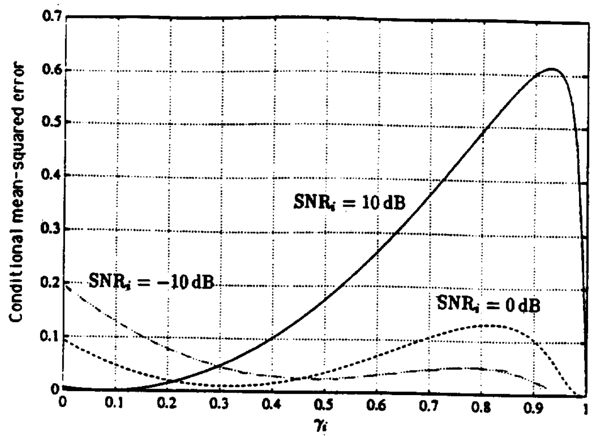


Fig. 8. Conditional MSE for the Bayes estimates of γ_i for per-mode SNR's of −10, 0, and 10 dB.

0, and 10 dB. Notice how the probability mass is concentrated near $\gamma_i = 0$ for the −10 dB curve and near $\gamma_i = 1$ for the 10 dB curve. There is a radical shift in the probability mass for a change in per-mode SNR of only 20 dB. It is not uncommon for least-squares problems to have per-mode SNR's that range over 60 to 80 dB. Some of the γ_i densities for these SNR's would show extreme probability mass concentrations near $\gamma_i = 0$ or $\gamma_i = 1$. In other words, true γ_i coefficients away from zero are very unlikely for modes with low SNR's, yet the constrained unbiased estimator is quite likely to produce estimates of γ_i spread over the range [0, 1] unless the per-mode SNR is quite large. When γ_i should be close to zero, the constrained unbiased estimator is likely to return a value far different from zero. Consequently, the solution can be very inaccurate because a small singular value does not get sufficiently damped.

C. Companders and Performance

Using the results of the appendix, we can analyze the performance of the conditional mean estimator, which defines a compander from z_i to $[\hat{\gamma}_i]_{\text{CM}}$. The conditional mean and conditional MSE are

$$E[[\hat{\gamma}_i]_{\text{CM}} \mid \gamma_i] = E[E[\gamma_i \mid z_i] \mid \gamma_i] \tag{5.8}$$
$$E[([\hat{\gamma}_i]_{\text{CM}} - \gamma_i)^2 \mid \gamma_i] = E[(E[\gamma_i \mid z_i] - \gamma_i)^2 \mid \gamma_i].$$

These functions have been computed numerically and plotted in Figs. 7 and 8. As before, the three curves in each figure correspond to per-mode SNR's of −10, 0, and 10 dB. Fig. 7 shows that, for small per-mode SNR's, the estimators are strongly biased toward the $[\gamma_i]_W$ value which parameterizes each of the curves. Shown in Fig. 8 is the conditional MSE for each of the chosen estimators.

VI. Simulations

In order to test the practicality of our results, we have applied them to several synthetic inversion and filtering problems, both overdetermined and underdetermined. After picking a suitable model matrix $\mathbf{H}$, we picked the parameter vector $\underline{\theta}$ using a random number generator and then computed the

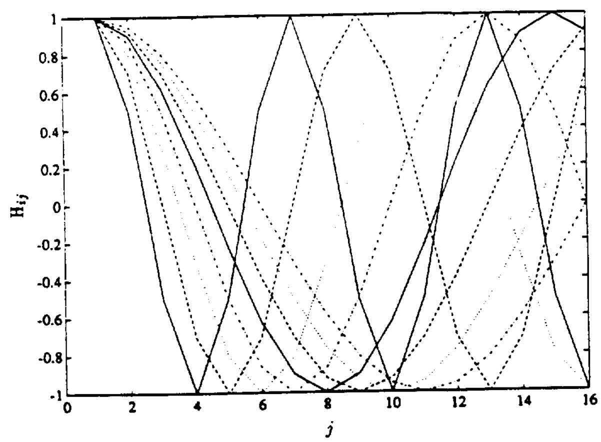


Fig. 9. Columns of the model matrix **H** for the overdetermined simulation.

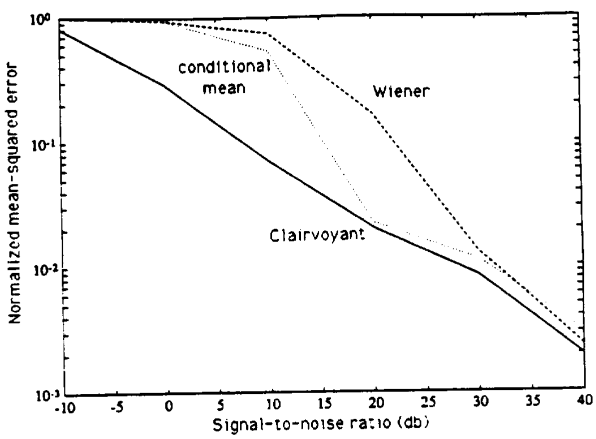


Fig. 10. Comparison of the MSE for $\underline{\boldsymbol{\theta}}$ for atypical realizations of $\underline{\boldsymbol{\theta}}$ in the overdetermined case.

signal $\mathbf{x} = \mathbf{H}\underline{\boldsymbol{\theta}}$. Different observations **y** were formed from the signal **x** by adding multiple noise realizations **n**, also picked with a random number generator. The rank-shaped estimates $\hat{\underline{\boldsymbol{\theta}}}_r$ and $\hat{\mathbf{x}}_r$ were computed and then compared to the true $\underline{\boldsymbol{\theta}}$ and **x**. Finally, the mean-squared errors in the solutions were averaged and plotted.

H was chosen to be 16×8 for the overdetermined case and 8×16 for the underdetermined case. We chose the columns of **H** to be discrete cosines with closely spaced frequencies so that **H** would be moderately ill-conditioned. The frequency of the ith column was picked to be $\pi/i + 2$. These columns are plotted in Fig. 9 for the overdetermined case.

To explore the merits of the adaptability of the $[\hat{\gamma}_i]_{\mathrm{CM}}$, we ran an overdetermined inversion simulation in which an atypical $\underline{\boldsymbol{\theta}}$ was picked. The power in the fourth mode of the signal was decreased by a factor of ten and that in the sixth and eighth modes increased by a factor of ten. The MSE for the simulation is plotted in Fig. 10. The data-adaptive $[\hat{\gamma}_i]_{\mathrm{CM}}$ method outperforms the Wiener solution at several SNR's. Consequently, the nonlinear, data-dependent $[\hat{\gamma}_i]_{\mathrm{CM}}$ method is a good alternative to the Wiener solution when atypical realizations of $\underline{\boldsymbol{\theta}}$ can produce atypical data.

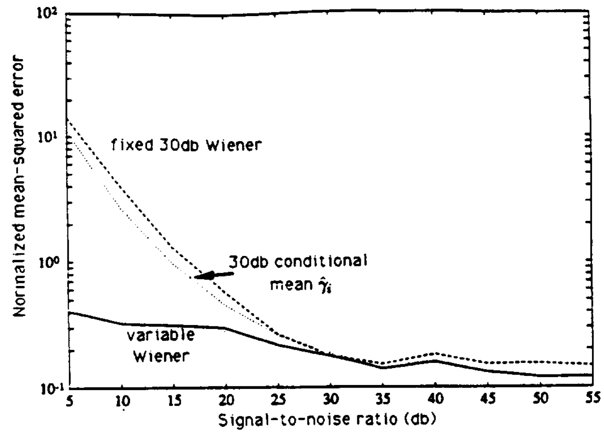


Fig. 11. Performance of the Wiener filter and $[\hat{\gamma}_i]_{\mathrm{CM}}$ filter when the true SNR is varied from the assumed SNR.

To further illustrate the adaptability of the $[\hat{\gamma}_i]_{\mathrm{CM}}$ for rank shaping, we ran a simulation to test the sensitivity of each estimator to the assumed signal power σ_S^2. We studied the overdetermined inversion problem using the previously described model matrix **H**, and the nominal SNR was chosen to be 30 dB. Multiple signal realizations were generated at different signal powers to yield SNR's from 5 dB to 55 dB, and the estimates of $\underline{\boldsymbol{\theta}}$ and **x** were computed. Three estimators were considered. The first estimator was a Wiener filter whose parameters changed to match the actual SNR for each realization of the observation **y**. The second estimator was a fixed Wiener filter designed for the nominal SNR of 30 dB. The third and final estimator used conditional mean estimates of γ_i in its inversion solution and was also designed for a SNR of 30 dB. The MSE for $\underline{\boldsymbol{\theta}}$ is shown in Fig. 11. The figure shows that the variable Wiener filter, which is the minimum MSE estimator for each SNR, bounds the error for the other estimators. The fixed Wiener filter performs identically to the variable Wiener filter at the designed SNR of 30 dB, but its relative performance degrades at other SNR's, as might be expected. The $[\hat{\gamma}_i]_{\mathrm{CM}}$ rank-shaped method, however, is data adaptive. It performs only marginally worse than the Wiener filter at 30 dB and performs a few dB better than does the fixed Wiener filter at SNR's more than 5 dB lower than the nominal SNR. These results indicate that the nonlinear rank-shaped estimator, because it is data dependent, has a performance advantage over the fixed Wiener filter in situations where the signal power is not precisely known or where the signal power varies between realizations.

VII. Conclusion

We have developed procedures for computing rank-shaped solutions to inversion and filtering problems. The rank-shaped estimators use weighting coefficients that depend on the data and a prior model. That is, the rank-shaping weights do not depend exclusively on the prior model as in other SVD-based methods. We have shown that rank shaping for the problems described is equivalent to estimating SNR's in modes of the signal. We have developed four data-dependent estimates of the clairvoyant weighting coefficients γ_i: $\hat{\gamma}_i$, $[\hat{\gamma}_i]_A$, $[\hat{\gamma}_i]_{\mathrm{ML}}$, and

$[\hat{\gamma}_i]_{\text{CM}}$. The $[\hat{\gamma}_i]_{\text{CM}}$ method, which assumes some prior information in order to bootstrap a solution, performs only slightly worse than the Wiener solution when data is typical and does much better when the data is atypical. The Wiener solution gets its minimum MSE property by performing well on typical realizations and poorly on atypical ones. The conditional-mean rank-shaping solution only slightly underperforms the Wiener solution for typical realizations and dramatically improves on it for atypical realizations.

The form of the $[\hat{\gamma}_i]_{\text{CM}}$ solution can be summarized as follows:

$$\hat{\underline{\theta}} = \mathbf{V}E[\underline{\Gamma} \mid \mathbf{y}]\underline{\Sigma}^{-1}\mathbf{U}^T\mathbf{y}; \qquad \hat{\mathbf{x}} = \mathbf{U}E[\underline{\Gamma} \mid \mathbf{y}]\mathbf{U}^T\mathbf{y}. \quad (7.1)$$

The matrix $\hat{\Gamma} = E[\Gamma \mid \mathbf{y}]$ is determined from the scalar companding curves of Fig. 5, which use the sufficient statistics $z_i = \hat{\beta}_i^2/(1+\hat{\beta}_i^2)$, where $\hat{\beta}_i^2 = (\mathbf{u}_i^T\mathbf{y}/\sigma)^2$ is a coarse estimate of SNR in mode i. These solutions are of a similar form to the Wiener solutions but use nonlinear functions of the data to determine the mode weights $\hat{\gamma}_i$ in the data-dependent matrix $\hat{\underline{\Gamma}} = \text{diag}(\hat{\gamma}_i)$. Our results are a logical extension of linear estimators to nonlinear estimators. They show that naive replacement of clairvoyant or Wiener coefficients with plausible estimates of them does not work. Something more sophisticated like conditional mean estimates of weighted coefficients, which leads to *very* conservative rank shaping, is required.

Appendix
The Conditional Density of γ_i Given z_i

From Bayes rule, we have

$$f(\gamma_i \mid z_i) = \frac{f(z_i \mid \gamma_i)f(\gamma_i)}{f(z_i)}. \quad (1)$$

In this simplified notation, we use $f(\gamma_i \mid z_i)$ as the conditional density for γ_i, given z_i, rather than $f_{\gamma_i|z_i}(t \mid s)$. We must compute each of the three densities on the right-hand side of (1). Let us define a few functions to make this task easier.

$$\nu(t) = \frac{t^2}{t^2+1}; \qquad \mu(t) = \sqrt{\frac{t}{1-t}} \quad (2)$$

$$\rho(t) = \frac{1}{\sqrt{2\pi}}\frac{d}{dt}\mu(t) = \frac{1}{\sqrt{2\pi}}\frac{\mu(t)^2}{2t^2}. \quad (3)$$

In order to avoid needless bookkeeping in the following derivations, the variables γ_i and z_i will be assumed to have values only in the range [0, 1]. Then $\mu(t)$ and $\nu(t)$ will be inverses of each other.

We will first compute the density of z_i given γ_i. We can write the distribution function as

$$F_{z_i|\gamma_i}(t, |, \gamma_i) = P[z_i \le t \mid \gamma_i] = P\left[\hat{\beta}_i^2 \le \frac{t}{1-t}\middle|\gamma_i\right]. \quad (4)$$

We know that $\hat{\beta}_i \mid \beta_i$ is distributed as $N[\beta_i, 1]$. This means that the random variable $\hat{\beta}_i^2 \mid \beta_i$ is invariant to the sign of β_i. Consequently, we can write $\beta_i = \pm\mu(\gamma_i)$, and (4) can be reduced to

$$\begin{aligned} F_{z_i|\gamma_i}(t \mid \gamma_i) &= P[-\mu(t) - \mu(\gamma_i) \le \hat{\beta}_i - \mu(\gamma_i) \le \mu(t) - \mu(\gamma_i) \mid \gamma_i] \\ &= \Phi(\mu(t) - \mu(\gamma_i)) - \Phi(-\mu(t) - \mu(\gamma_i)) \end{aligned} \quad (5)$$

where $\Phi(x)$ is the integral of the normal density function. The density is computed by differentiating the distribution

$$f(z_i \mid \gamma_i) = \rho(z_i)\left[\exp\left\{-\frac{c_0}{2}\right\} + \exp\left\{-\frac{c_1}{2}\right\}\right] \quad (6)$$

where c_0 and c_1 are defined to be

$$c_0 = (\mu(z_i) - \mu(\gamma_i))^2 \quad (7)$$

$$c_1 = (\mu(z_i) + \mu(\gamma_i))^2. \quad (8)$$

Now, let us compute the unconditional density for γ_i using the above technique

$$\begin{aligned} F_{\gamma_i}(t) &= P[\gamma_i \le t] \\ &= P[-\mu(t) \le \beta_i \le \mu(t)] \end{aligned} \quad (9)$$

where β_i is distributed as $N[0, (\sigma^{-1}\sigma_S\sigma_i)^2]$. Define the variable a to be the square root of the SNR

$$a = \frac{\sigma_i\sigma_S}{\sigma} \quad (10)$$

and substitute into (19) to get

$$\begin{aligned} F_{\gamma_i}(t) &= P\left[-\frac{\mu(t)}{a} \le \frac{\beta_i}{a} \le \frac{\mu(t)}{a}\right] \\ &= \Phi\left(\frac{\mu(t)}{a}\right) - \Phi\left(-\frac{\mu(t)}{a}\right) \\ &= 2\Phi\left(\frac{\mu(t)}{a}\right) - 1. \end{aligned} \quad (11)$$

The density function for γ_i is then the derivative of $F_{\gamma_i}(t)$ evaluated at γ_i

$$f(\gamma_i) = \frac{2\rho(\gamma_i)}{a}\exp\left\{-\frac{c_2}{2}\right\} \quad (12)$$

where c_2 is defined to be

$$c_2 = \left(\frac{\mu(\gamma_i)}{a}\right)^2. \quad (13)$$

We compute the density of z_i by integrating the joint density for z_i and γ_i over all γ_i. Using the results of (6) and (13), we get

$$\begin{aligned} f(z_i) &= \int_{-\infty}^{\infty} f(z_i \mid \gamma_i)f(\gamma_i)\,d\gamma_i \\ &= \int_0^1 \rho(z_i)\left[\exp\left\{-\frac{c_0}{2}\right\} + \exp\left\{-\frac{c_1}{2}\right\}\right] \\ &\quad \times \frac{2\rho(\gamma_i)}{a}\left[\exp\left\{-\frac{c_2}{2}\right\}\right] d\gamma_i. \end{aligned} \quad (14)$$

Define the variables

$$b = \nu(a)$$

$$c_3 = \frac{1}{b}[(\mu(\gamma_i) - b\mu(z_i))^2] + (1-b)\mu(\gamma_i)^2$$
$$c_4 = \frac{1}{b}[(\mu(\gamma_i) + b\mu(z_i))^2] + (1-b)\mu(\gamma_i)^2$$
$$c_5 = \frac{\mu(z_i)}{a^2+1}; \qquad d_0 = c_3 - c_5; \qquad d_1 = c_4 - c_5 \quad (15)$$

and then the square in each exponent is completed. Equation (14) reduces to

$$f(z_i) = \rho(z_i)\exp\left\{-\frac{c_5}{2}\right\} \times \int_0^1 \frac{2\rho(\gamma_i)}{a}\left[\exp\left\{-\frac{d_0}{2}\right\} + \exp\left\{-\frac{d_1}{2}\right\}\right] d\gamma_i. \quad (16)$$

Now, define t to be

$$t = \mu(\gamma_i); \qquad dt = \frac{\mu(\gamma_i)^3}{2\gamma_i^2}\, d\gamma_i = \sqrt{2\pi}\rho(\gamma_i)\, d\gamma_i, \quad (17)$$

so that (16) becomes

$$f(z_i) = \rho(\gamma_i)\exp\left\{-\frac{c_5}{2}\right\}\frac{2\sqrt{b}}{a} \times \int_0^\infty \frac{1}{\sqrt{2\pi b}}\left[\exp\left\{-\frac{d_2}{2b}\right\} + \exp\left\{-\frac{d_3}{2b}\right\}\right] d\gamma_i \quad (18)$$

where d_2 and d_3 are

$$d_2 = (t - b\mu(z_i))^2; \qquad d_3 = (t + b\mu(z_i))^2. \quad (19)$$

By simplifying the integral, we get the desired density for z_i

$$\begin{aligned} f(z_i) &= \rho(z_i)\exp\left\{-\frac{c_5}{2}\right\}\frac{2\sqrt{b}}{a}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi b}}\exp\left\{-\frac{d_2}{2b}\right\} d\gamma_i \\ &= \rho(z_i)\exp\left\{-\frac{c_5}{2}\right\}\frac{2\sqrt{b}}{a} \\ &= \frac{2\rho(z_i)}{\sqrt{a^2+1}}\exp\left\{-\frac{\mu(z_i)^2}{2(a^2+1)}\right\}. \end{aligned} \quad (20)$$

Substitute the results of (6), (12), and (20) into (1) and simplify to get the final desired result

$$f(\gamma_i \mid z_i) = \frac{\rho(\gamma_i)}{\sqrt{b}}\left[\exp\left\{-\frac{e_0}{2}\right\} + \exp\left\{-\frac{e_1}{2}\right\}\right] \quad (21)$$

where

$$e_0 = \frac{1}{b}(\mu(\gamma_i) - b\mu(z_i))^2; \quad e_1 = \frac{1}{b}(\mu(\gamma_i) + b\mu(z_i))^2$$
$$b = [\gamma_i]_W = \frac{\sigma_S^2\sigma_i^2}{\sigma_S^2\sigma_i^2 + \sigma^2}. \quad (22)$$

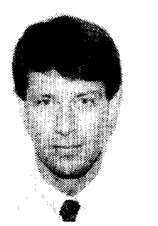
The parameter b, which is the Wiener gain $[\gamma_i]_W$, completely parameterizes this conditional density.

ACKNOWLEDGMENT

The authors acknowledge an anonymous reviewer who suggested the term "rank shaping" as a better description of our method than the term "rank reduction" which we had originally proposed. We thank K. Duell for pointing out to us the work of James, Stein, and others on shrinkage methods.

REFERENCES

[1] W. James and C. Stein, "Estimation with quadratic loss," in *Proc. 4th Berkeley Symp. Math. Statist. Probability*, vol. I, 1961, pp. 361–379.
[2] D. W. Marquardt, "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation," *Technometrics*, vol. 12, no. 3, pp. 591–612, 1970.
[3] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, 1981.
[4] T. S. Huang and P. M. Narendra, "Image restoration by singular value decomposition," *Appl. Opt.*, vol. 14, no. 9, Nov. 1975.
[5] B. R. Hunt, "A theorem on the difficulty of numerical deconvolution," *IEEE Trans. Audio Electroacoust.*, vol. 20, no. 1, pp. 94–95, Mar. 1972.
[6] C. L. Mallows, "Some comments on C_p," *Technometrics*, vol. 15, no. 4, Nov. 1974.
[7] J. A. Scales, P. Docherty, and A. Gersztenkorn, "Regularisation of nonlinear inverse problems: Imaging the near-surface weathering layer," *Inverse Problems*, vol. 6, no. 1, pp. 115–131, Feb. 1990.
[8] L. L. Scharf, "Topics in statistical signal processing," Course 2 in *Signal Processing*, J. L. Lecoume, T. S. Durrani, and R. Stora, Eds. Amsterdam, The Netherlands: Elsevier, 1987.
[9] ——, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
[10] L. L. Scharf and J. Storey, "Rank reduction and order selection for parametric spectrum models," in *Proc. 3rd ASSP Workshop Spectrum Anal.*, Boston, Nov. 1986.
[11] A. J. Thorpe, "Reduced rank methods for solving least squares problems," M.S. thesis, Univ. of Colorado, Boulder, May 1990.
[12] A. J. Thorpe and L. L. Scharf, "Reduced-rank methods for solving least squares problems," in *Proc. 23rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1989, pp. 609–613.



Anthony J. Thorpe holds an M.S. in electrical engineering and a B.S. in aerospace engineering, both from the University of Colorado.

He is currently Director of Technical Services at Analytical Surveys, Inc., a company specializing in the conversion of data for Geographical Information Systems (GIS). His responsibilities include the coordination of research, software development, technical training, and project initiation.

Louis L. Scharf (S'67–M'69–SM'77–F'86) received the Ph.D. degree in electrical engineering from the University of Washington, Seattle, in 1969.

From 1969 to 1971, he was a member of the Technical Staff at Honeywell's Marine Systems Center in Seattle. He was a faculty member at Colorado State University, Fort Collins, from 1971 to 1981, where he last served as a Professor of Electrical Engineering and Statistics. From 1982 to 1985, he was a Professor and Chair of Electrical Engineering at the University of Rhode Island, Kingston. He is currently a Professor of Electrical and Computer Engineering at the University of Colorado, Boulder, where he teaches and conducts research in signal processing. In 1974, he was a Visiting Associate Professor at Duke University, Durham, NC. In 1977, he was at the University of South Paris, Orsay, where he was a member of the Technical Staff in the CNRS Laboratoire des Signaux et Systemes, Gif-sur-Yvette. In 1981, he was a Visiting Professor at the University of La Plata, Argentina, and a Visitng Professor at the Ecole Nationale Supérieure des Télécommunications, Paris. He has served as a consultant to Honeywell Inc., the Applied Physics Labs, Seattle, the Research Triangle Institute, Green Mountain Geophysics, and Ball Aerospace Corporation.

Dr. Scharf is a past member of the ASSP ADCOM and the Editorial Board of *Signal Processing*. He is a past Associate Editor for the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. He was the Technical Program Chairman for the 1980 International Conference on ASSP. He was appointed IEEE National Distinguished Lecturer for 1993–1994 and is a member of Eta Kappa Nu.